a) In all cases, we plan close cooperation with the users in all aspects of the problem. Although the basic isolation procedures are the problem of each investigator, his knowledge of the available facilities and their limitations can be an important aid to sample preparation and analysis of the results. This is particularly true for collaborators who are unfamiliar with the techniques of HRMS (e.g., sample size and resolving power necessary to separate the mass doublets that can be realistically expected in different contexts).

b) The needs of the user community will be varied. Drs. Duffield and Smith will, in collaboration with the users, determine the kinds of MS experiments which will be most useful, considering sample complexity, stability, quantity, and so forth. We wish to utilize fully the existing resource and our proposed extensions, bringing to bear on a problem any technique which is appropriate and can be provided. This will include the full scope of available experimental techniques in MS (LRMS, HRMS, GC/LRMS, GC/HRMS, metastable defocussing, ultra-high resolution mass measurements) and available computer programs (see below).

c) Many problems will be amenable to treatment by computer programs which exist or which will be developed, for example, structural isomer problems or HRMS interpretation on compounds in a well-understood class. We will take the responsibility for utilizing these programs where appropriate to assist in structure elucidation problems. We will instruct members of the community in use of the programs when programs are used routinely by collaborators.

III. METHODS

Molecular structure elucidation entails the intelligent and patient application of a large body of knowledge to each specific problem. The importance and relative difficulty of the problem impel us to seek the powerful assistance of computer programs to help chemists in their analyses. It is unlikely that such programs will ever replace chemists, especially because computer programs are readily written only to focus on rather narrow aspects of problems. Nevertheless, our past research is reasonably forwarded as a demonstration of the computer's ability to assist in practical biomolecular characterization although this was a spinoff from theoretically oriented research.

In order to meet the major objectives of this proposal we will focus our attention primarily on structure elucidation of biomedically important compounds through MS and AI. However, many of the computer programs can already use information from other analytical techniques. So we want to be able to think of structure elucidation in the context of an ensemble of analytical capabilities.

A. Enhancing the Power of the Mass Spectrometry Resource

We have developed a significant resource consisting of instrumentation (the Varian MAT-711 and ancillary equipment) and computer programs for instrument evaluation, data acquisition and reduction. Routine reduction of high resolution mass spectra to elemental compositions and ion abundances without human intervention provides the capability for efficient handling of large volumes of high resolution mass spectra (such as will result from GC/HRMS runs). The development of the GC and of the GC/MS combination is in the excellent hands of Ms. Annemarie Wegmann, who is responsible for operation of the complete system. We now have more than two years of operational experience with the MS, the GC and related equipment under a wide variety of experimental conditions.

None of the resource-related research discussed in this proposal can be carried out without significant quantities of mass spectral data. The existence and extensions of the MS resource, the development of computer techniques and the applications to biomedical problems demand an efficient mechanism for acquisition and reduction of MS data, and eventual transmission of the data to the SUMEX resource. Thus, operation of the MS requires substantial computer support to deal with the large volumes of data produced by the system at high data rates. We feel that a properly configured system of hardware and software should provide, at a minimum, the following capabilities:

1) Detailed evaluation of the condition and performance of the MS prior to recording data on valuable samples, with feedback to the operator.

2) A coordinated system of hardware and software for signal conditioning, peak detection and peak analysis.

3) Data reduction techniques based on a computed (not theoretical) model of the MS, including peak shapes, mass/time function, and resolving power as a function of mass.

4) Peak profile analysis for multiplet detection and resolution.

5) Computer control of scan rates, clock rates and optimum analog and digital filtering parameters.

6) Some on-line feedback, to the operator, assessing the performance of the system during an experiment.

7) The system must deal with frequent or repetitive HRMS scans, requiring the capability for rapid storage and analysis of large volumes of data.

Previous support of our research by the NIH and NASA has given us a firm foundation of programs and experience. We have, up until the termination of the ACME computer facility (July 31, 1973), demonstrated capabilities 1-5 above. We were precluded from pursuing capabilities 6 and 7 due to the configuration of the ACME facility.

The demise of the ACME computer facility and the subsequent incorporation of the PL/ACME language into a new IBM 370/158 facility under Stanford auspices has forced a reevaluation of the means for providing HRMS laboratory computing support. We had previously depended exclusively on ACME for data reduction processing. The ACME transition poses both technical and fiscal decisions in that the real-time support capabilities of the new facility will be different from ACME's and the fee for service basis of the facility requires an explicit budget allocation for its use. Previously we had received ACME computing support without charge as part of the core research effort. Since we were thus required to revise our computing plans, we have explored a number of options for near-term as well as longer term solutions.

As outlined in the attached annual report, we have chosen an interim approach (through the end of the current grant, 4/30/74) which minimizes near-term costs, including hardware and software conversions as well as operating expenses. This approach entails connecting the MAT-711 spectrometer to the 370/158 computer through an IBM 1800 interface. It allows use of the existing PL/ACME programs but will have real-time response limitations at least as severe as ACME had (which is inadequate for either GC/LRMS or GC/HRMS). Our existing computing budget provides for only a very low level of instrument utilization in this mode.

For a longer term solution these constraints are unacceptable. Current estimates are that continued use of the inadequate 1800-370/158 connection and PL/ACME interactive programs under full instrument productivity would cost up to $4,100 per month. Three alternatives have been investigated for improving technical performance and reducing cost. This review has resulted in our current proposal to augment the existing mini-computer system (PDP-11/20) with local storage and arithmetic capabilities. This stand-alone system would not support real-time, on-line data reduction but would allow routine data acquisition and instrument performance evaluation, followed by off-line data reduction.

Alternatives considered include:

1) Modified 370/158 Connection

We discussed with personnel in the Stanford Center for Information Processing (SCIP) various approaches for improving 370/158 service. Detailed planning is still under way within SCIP in regard to real-time support and future pricing policies. Thus the following conclusions are tentative. It appears that the long-term cost would be prohibitive to continue real-time data acquisition by the 370/158. Instead, a store-and-forward system was proposed. This would entail an augmentation of the existing PDP-11/20 front-end mini-computer with memory, disk, tape, and a new interface to the 370/158, totaling about $28,000. This approach is workable, if limited near real-time instrument performance evaluations could be made to assure satisfactory instrument setup and data acquisition. It was recommended that the existing software be converted from PL/ACME to a more efficient language (such as FORTRAN) to reduce operating costs. This would require approximately 4-6 man-months of effort. The resulting decrease in operating costs could not be estimated in time for this proposal because the new SCIP pricing policies are not formulated and inadequate 370/158 system analysis tools are operational to evaluate our benchmarks in terms of detailed resource consumption. We have therefore budgeted an approach based on the remaining two options with the understanding that we will reconsider the SCIP option before proceeding with an implementation should this proposal be funded.

2) SUMEX

The recently approved AIM-SUMEX PDP-10 facility will provide necessary computing support for the development and use of DENDRAL AI programs. The MS laboratory produces data which these programs analyze and thus has a close relationship to the AI research. The SUMEX computer could help in the off-line reduction of instrument data, particularly during the early stages of the project when the machine load will be relatively light (20-25%). The present programs would require conversion from PL/ACME as in option (1), which would take 4-6 man-months. Such computing may use from 15-30 minutes of CPU time per day, depending on the amount of GC/MS work. While this approach saves operating computing costs, the front-end PDP-11/20 would require augmentation as in option (1) ($28,000) to allow store-and-forward operation with subsequent off-line data reduction on SUMEX. This is needed because SUMEX is not configured to allow real-time acquisition of the volume of data anticipated. This approach, while the least costly, would entail a measurable use of the PDP-10 resources which we feel are better reserved for the intended AIM-SUMEX applications. In addition, because of the priorities anticipated for allocation of SUMEX to AI research, particularly as loading increases, scheduling may be required which will constrain the MS laboratory operation. For these reasons, we feel a better, even though slightly more expensive, approach is a stand-alone PDP-11/20 data reduction system.

3) Stand-Alone PDP-11/20

The augmentations of the existing front-end PDP-11/20 required for store-and-forward operation in conjunction with the 370/158 or with SUMEX, come close to meeting the needs of a stand-alone data system. In addition to the memory, disk and tape, an augmented arithmetic capability is needed to allow rapid floating point calculations. A special device for this purpose costs about $7,500. The SUMEX interface can be less sophisticated in this case, however, accounting for the much lower data volume after reduction, so that the total cost of the stand-alone system would be $34,000. As with the other options, conversion of the present programs would be required.

This approach, while slightly more expensive, has the advantages of off-loading all data logging and reduction functions from SUMEX and affords an adequate capacity for non-real-time, stand-alone data reduction on the PDP-11/20. It furthermore allows more freedom and responsiveness in the operation of the MS laboratory since data collection or reduction can be scheduled without worrying about the impact on AIM-SUMEX users. We therefore propose and have budgeted an augmentation of our existing mini-computer system as a stand-alone data reduction facility.

The biomedical community (see User Community, Sec. I.B.4 above) desiring access to our facilities for structure elucidation have a variety of problems, some of which can be solved by existing instrumentation and computer techniques, as noted above. However, many problems consist of complex mixtures of compounds where analysis by conventional GC/LRMS does not lead to unambiguous solutions, and separation of components on a preparative scale for other spectroscopic analysis is difficult (e.g., see marine sterols, section D, below). These problems are amenable to attack by a system comprised of a GC/HRMS combination, the GC providing separation, coupled with the MS operating at high resolution to provide elemental compositions. Thus, upgrading of our current system so that GC/HRMS data can be provided on a routine basis is a desirable, and we believe necessary, step to solve many of these problems.

We propose to continue the development of the GC/HRMS system while maintaining existing capabilities of routine HRMS analysis and GC/MS where this efficiently responds to local needs. Many members of the user community will require in addition to GC/HRMS, HRMS analysis of relatively pure compounds or mixtures of small numbers of compounds. We will provide this capability on an interim basis, using Stanford's IBM 370/158 system while the PDP 11/20 system is being upgraded.

We were able, using the ACME computer facility, to start evaluating the operation of a GC/MS system at high mass resolutions. These experiments were hampered somewhat by the limitations of the computer system used to acquire the data (only occasional, single scans were possible); they were necessarily discontinued (as well as all HRMS operation!) upon the termination of ACME. We do have, however, some benchmark figures for the performance of the proposed system. Mixtures of fatty acid esters (e.g., methyl palmitate and methyl stearate) gave good quality

mass measurements (+-10 ppm) over a dynamic range of 100:1 for sample sizes of the order of 0.5-1.0 micrograms/component during 10 sec/decade in mass scans (resolving powers 5,000-8,000).

We are haltingly continuing our evaluation of the GC/HRMS system even without a data system, making measurements on individual ions of the mass standard and known materials in the GC effluent. These data can be approximately translated into expectations during dynamic scanning. We have performed an extensive series of measurements on both methyl stearate and cholesterol (not derivatized), the latter compound being more representative of our current research problems. These measurements tend to confirm the preliminary data described above. Firmer data will be available subsequent to the submission of this proposal.

We propose to operate our existing GC/MS system under high resolution conditions aiming toward optimization of resolving powers, scan rates and GC and molecular separator operating conditions to determine the maximum usable sensitivity of the system.

We recognize that the ultimate sensitivity will not approach that attainable by photographic methods of recording; we feel that the ability for on-line operation and evaluation of the operating conditions of the MS partially offsets the sensitivity disadvantages. We realize that some structure elucidation problems will not be amenable to study because of the sensitivity limitations; we feel, however, that many problems of interest to the User Community can be studied effectively with this performance capability. Rather than propose a research program to increase the sensitivity of high resolution mass spectrometers (e.g., McLafferty, et.al., Anal. Chem., 44, 2282 (1972), dynamic rescanning of peaks; Jet Propulsion Laboratory - chemical multiplier emission/detector arrays, private communication to T. Rindfleisch), we propose to identify our limitations and, with our collaborators, use discretion in selecting and preparing samples.

Further accelerations of technical capability to meet the state of the art in sensitivity will require investments in hardware that can be better justified at a later stage of a successful facility program . Meanwhile, other laboratories can be expected to make significant contributions to this important problem. Practical regard for budget limitations is the main reason we do not press this issue ourselves at the present time.

Significant improvements in sensitivity (with only small decreases in mass measurement accuracy) can be achieved by operating the MS at reduced resolving powers coupled with intelligent analysis of the resulting data to detect and resolve the potentially greater number of overlapping peak envelopes. This proposal is not entirely new (e.g., see Smith, et.al., Anal. Chem., 43, 1796 (1971); Burlingame, et. al., in "Computers in Analytical Chemistry," C.H. Orr and J.A. Norris, Ed., Progress in Analytical Chemistry, Vol. 4, Plenum Press, New York, N.Y., 1970, Chap. III). We can, however, significantly extend these earlier techniques by utilization of our multiplet resolution algorithm. This algorithm embodied in a computer program, has been shown to increase the effective resolving power of the MS up to a factor of three. It bases its operation on a dynamic model of peak shape computed directly from the data. For computational efficiency and to avoid

spurious information, this algorithm would be best implemented as a post-processor, basing its search for multiplets on the results of prior elemental composition determination.

The ability to detect and analyze for unresolved peaks is mediated by consideration of the mass measurement accuracy of an MS system. These systems are capable of determining peak positions (and thus masses) to a small fraction of the peak width. The high accuracy of such measurements (+- 2-10 ppm) can, in fact, be utilized to detect and "resolve" multiplets in instances where the unresolved species are known precisely (see Burlingame, et al., ref. above, for CH vs. 13C doublet detection and resolution).

For instances where the heteroatom content of a molecule is known or where the possibilities are reduced severely by chemical, spectroscopic and mass measurement heuristics, there may be a range of possible overlapping ions resulting from fragmentation of the molecule. These potential overlaps may be computed and then used (in combination with the known resolving power and mass measurement accuracy of the MS and the measured mass of the peak, assuming it was comprised of only one type of ion) to direct the multiplet resolution program.

As an example, we have computed the possible mass doublets for various ranges of compositions (Lederberg, et al., to be published). A sample table for C, N, O =<4 is appended (Table 1). Only 28 of the 364 possibilities are shown, namely those whose mass difference (e) <.05 mass units. *

Of these 28, 13 show e>.03 and would be fairly easy to resolve, requiring 1/5000 resolution at MW=150.

At the other extreme, 5 doublets show e<.01 (CN4 vs. H4O4; C2H2O vs. N3; C2N2 vs H4O3; C3N vs H2O3; and C4 vs H2NO2) which would demand special treatment for resolution.

The 10 doublets for which .01 =< e =< .03 pose the interesting challenges for tradeoff of resolution vs. sensitivity in the context of given problems. For example, if N is absent, the only ambiguities are C3 vs. H4O2 (e = -.02) and C4 vs O3 (e = .015).

Much as we would wish always to have unambiguous empirical formulas for all ions, HRMS remains a valuable tool despite these limitations. As shown by these examples, even moderate resolution reduces the number of candidates to a manageably small number of alternatives. Contextual and interval data (within the spectrum) can be used to trim these further at two levels: (a) pooling of peak statistics to sharpen decision probabilities on the presence of heteroatoms -- the fragments are subsets of the molecule and (b) the assemblage of candidate solutions under each of the alternative formulae. Manifestly, computer processing can sort out branches of decision trees that would soon exhaust human patience.

These heuristics are built into the DENDRAL programs (solutions based on fragmentation theory), but are also applicable to table look-up approaches.

We (ref. 28,33), and others (e.g., H.-K, Wipf, et. al., J. Amer. Chem. Soc., 95, 3369 (1973)) have illustrated the importance of

metastable ion determinations in automated structure elucidation based on MS data. Data on metastable ions must be judiciously selected because of the time and sample normally required to perform the measurements. Our programs are now capable of precise specification of those experiments necessary and sufficient to distinguish among a set of candidate structures. We seek more efficient ways of acquiring these selected instrumental data. This can be accomplished with minimal cost by developing the hardware and software necessary to perform (defocussed) metastable scans and calculate the data. Much of the hardware, except an accurate sensor for accelerating voltage, already exists. We have had considerable experience in peak detection on the software side; the calculations to determine transitions are simple. It is assumed that the operator would manually adjust the instrument to the desired "daughter" mass prior to initiation of the scan of metastable origins ("parents") of this daughter.

The recent availability of reversed-geometry instruments has provided new methods of metastable defocussing (e.g., Beynon, et.al., Anal. Chem., 45 (12), 1023A (1973)). We have illustrated the power of these techniques in mixture analysis (ref. 69). No "normal" geometry instrument is equipped to perform these measurements to determine all the daughters of a given parent, information which is frequently more useful than the converse. This information can be obtained, in principle, by synchronous variation of two of the three fields (magnetic/ accelerating/ electrostatic deflection) in a very accurate way. We would like to explore this possibility because we feel that this technique, if feasible, would represent a significant upgrading of the many standard geometry, double-focussing instruments in existence.

B. Computer Assisted Structure Elucidation

As mentioned above, some existing programs can be used immediately for structure elucidation problems using MS data. The programs have been described in detail elsewhere and are mentioned in the section on existing capabilities (Sec. I.B.E, above). The Planner's performance, for example, is excellent precisely in the areas where MS, by itself, is capable of definitive structure analysis. The general intellectual flexibility of the human chemist is beyond the reach of plausible programs. On the other hand, where the history of a sample is known, so as to restrict the potential classes of compounds and for classes where the rules of MS fragmentation are well understood, the program's performance matches that of trained mass spectroscopists, the program also offers some advantages in its exhaustive and rapid analysis of the data. Many structure elucidation problems of the user community fit into this category and existing resources can fulfill these needs.

Whether man- or computer-implemented, MS cannot solve all structure elucidation problems, however. In such cases, recourse is to other spectroscopic techniques if sample size permits. As described in the introductory section, diverse information is pieced together to achieve a solution. Interactive computer programs can assist in segments of this procedure, with the advantages of exhaustive evaluation of the data and the molecular structures suggested by these data.

In our own and in planned collaborative work, we will call upon the extensive facilities of the chemistry department for acquisition of additional spectroscopic data. These services are financed by fees, paid from existing research grants of the user community. There are sufficient documented examples of structure elucidation problems to obviate the requirement for extensive use of these additional facilities in development of the programs. On the other hand, the intensive pursuit of mechanized "intelligence" in the domain of MS requires more than availability of public MS data . It requires the collaboration of skilled chemists actively engaged in practical MS research and, at the same time, committed to the exploration of innovations in the application of AI to the solution of the problems

As in the past, we will develop the computer programs through close collaboration among Drs. Duffield and Smith (and other members of their groups) and the program designers and programmers. For us, this means daily consultation for discussion of strategy, extensions to the program, and solutions to new problems. In particular, we propose to continue software development (on the AIM-SUMEX facility) as folllows:

1) The recently completed structure generating algorithm will be the core of our efforts to assist in structure elucidation. The structure generator can guarantee that the correct solution is somewhere in the list of possibilities. Additional programs, such as the Planner allow us to avoid exhaustive generation in practice. Some parts of the cyclic structure generator program have not been extensively tested yet, and these tests will be the first task to complete.

2) The structure elucidation task is strongly directed toward rejection of whole categories (e.g., compound classes) of solutions as quickly as possible by using as much knowledge about the chemical history or characteristics of a sample as is available. Details of spectroscopic data then define the molecular framework more precisely. Each step in this procedure represents the application of constraints on the set of possible solutions. Computational efficiency demands that these constraints be applied early in the generation process when the structure generator is utilized.

We have made some effort to examine the kinds of constraints used by scientists engaged in structure elucidation. We have begun designing strategies so that these constraints can be brought to bear on the structure generator. Some of these strategies involve minor changes to the existing program; others require significant extensions of existing generating functions. One approach which seems particularly attractive to us is presently under development. This approach will utilize the existing structure generator, with some modifications, to generate a dictionary of cyclic skeletons up to those containing a maximum of twelve tertiary vertices. The dictionary will be a complete, irredundant list of ring systems which contain no multiple bonds and no cut-edges (acyclic parts). This dictionary will be organized and keyed so that many constraints can be implemented easily. The dictionary will allow exhaustive specification of ring systems with double bonds and/or aromaticity. The rings themselves can be labelled with heteroatoms to generate heterocyclic ring systems,

or with acyclic radicals to generate substituted ring systems. The existence of the dictionary will lead to greater computational efficiency as it needs to be generated only once, and specific configurations of rings (numbers, sizes, fusions) can be pulled immediately from the dictionary.

We propose to continue these investigations so that a reasonable variety of constraints can be recognized and utilized effectively by a computer program. This represents the first step toward increasing the chemical knowledge of a program which views molecular structures and their manipulation as mathematical entities and transforms.

3) Present, effective use of the structure generator or its subroutines for special problems requires a detailed knowledge of the program. We propose to develop an interface between users and the program to remove this requirement. The interface would contain elements of structure input and display routines and a simple language for application of constraints. Portions of these elements are available from other workers (e.g., Richard Feldman, NIH) and we would draw on these sources whenever possible.

4) We propose that initial efforts will be directed toward a system where the scientist examines his own data and inputs his findings (in terms of allowed and disallowed structural features) to the program as constraints. The generator would then provide a list of possible solutions to be evaluated, followed by iteration on this procedure.

5) Many structure elucidation problems can be characterized as assembly of sub-structures inferred from spectroscopic data into complete molecular structures. Although there are two instances in the literature describing programs with the capability to solve this problem (see S. Sasaki, "Determination of Organic Structures by Physical Methods, Vol. 5," F.C. Nachod and J.J. Zucherman, Ed., Academic Press, New York and London, 1973, p. 285; M.E. Munk, C.S. Sandano, R.L. McLean, and T.H. Haskel, J. Amer. Chem. Soc., 89, 4158 (1967)), we do not feel these approaches fulfill the requirements for generating complete lists of structures and avoiding duplicate structures. We have some strategies to solve this problem, thus extending the scope of the generator while tying it more closely to the methods used by chemists engaged in structure elucidation. Our existing structure generator has this capability; as long as the sub-structures are connected only by a single bond, no new rings are formed.

6) We wish to implement general routines for finding molecular ions from spectroscopic data in order to improve the general power of the Planning program. The current Planning program depends on having some metastable ion information with HRMS data, together with knowledge of the structural class with special rules for the class. We will incorporate strategies suggested by Biemann (K. Biemann and W.J. McMurray, Tet. Lett., 647 (1965)) and McLafferty (R. Venkataraghavan, F. W. McLafferty, and G. E. Van Lear, Org. Mass Spectrom., 2, 1 (1969)) for finding molecular ions, but also give the program the flexibility to use class-specific information when available. The procedure will be to use these kinds of information within a general heuristic search paradigm.

7) The section on aims indicated some longer-term directions

which might be pursued. Of these, we feel that the incorporation of three-dimensional information into the program is perhaps most important (e.g., representation of three dimensional information, molecular modelling including steric factors). Lederberg has previously discussed ways (Ref. 1) in which three dimensional information can be considered in the generation and representation of molecular structures. More recently, the work of Wipke (J. Amer. Chem. Soc, in press; personal communication) in connection with computer assisted organic synthesis has provided important results which we would attempt to utilize to avoid unnecessary duplication of effort. We plan to collaborate with Dr. G. Loew (Stanford Genetics Dept.) to utilize her available programs on molecular orbital methods to determine local minima for conformations.

Another longer term goal which we feel is both interesting and important is the use of an extended Predictor (which we have previously described in the context of MS) to assist in distinguishing among potential solutions to a structure elucidation problem. We have recently carried out some extensions to the existing Predictor by incorporating the ability to suggest metastable defocussing experiments. Further extensions to include knowledge about other spectroscopic techniques and the information which can be elicited from these techniques are clearly feasible and could be a powerful extension to our computer assistance efforts.

C. Theory Formation

One important aim of this project is to improve the existing theory formation capabilities and thus provide more assistance to scientists investigating regularities within classes of compounds. This is a theory formation task at a very pragmatic level. The MS theory that the program attempts to find is of the same form as the one practicing mass spectroscopists use for structure elucidation. Thus, resulting pieces of theory are extensions to both the scientists' theory and the computer's theory of the discipline. To improve this program we need to complete the Plan-Generate-Test program that has been started (as described in the appended annual report) and tune it over many test cases. We also wish to make the programs interactive and easy to use so that they are more readily accessible. This can be done when the programs are transferred to the AIM-SUMEX facility.

We plan to apply the theory formation program to two different kinds of data: (a) the data collected in the interest of understanding the mass spectrometry of a particular class of compounds, as was done for estrogenic steroids, and (b) collections of diverse data that may provide some insight into more general fragmentation mechanisms. For example, we hope to find general rules analogous to the alpha-cleavage rule or the stability of aromatic rings.

The INTSUM program mentioned in Section (I) is the planning phase of the theory formation program. It currently runs in batch mode on Stanford's 360/67 computer. We wish to add an interactive monitor to INTSUM to give an investigator the ability to set up his own conditions for interpreting the mass spectra and to control the type of summary he wishes to see. For example, if he

is interested in the allowable hydrogen transfers associated with one specific process the program could be instructed to produce a very specific summary. Also, we wish to add an interactive program for answering questions about the results. For example, an investigator should be able to find out easily how many processes involve cleavage of a specific bond and how strong their resulting MS peaks are.

The INTSUM program is now used routinely by mass spectroscopists at Stanford engaged in investigations of the mass spectrometric fragmentation of various classes of organic compounds, primarily steroids. A manuscript is now in preparation (Ref. 54) describing the fragmentation of progesterone and related compounds. The program was used extensively in this work. We are now beginning a detailed examination of the fragmentation of steroids related to the androstane skeleton, particularly the biologically important testosterones. We propose to continue to use the INTSUM program in its present form and as it is improved in support of these studies.

The generator of rules that we now have, RULEGEN, does a credible job of explaining the regularities summarized by INTSUM. It has found, for example, the well-known alpha-cleavage fragmentation process and beta cleavage followed by rearrangement in the low resolution data for fifteen aliphatic amines. The program will be extended in two important ways to increase its utility: (i) the program needs to be able to work with an increased number of descriptive predicates in the generation of rules, and (ii) it needs to be given a more flexible representation of complex fragmentation mechanisms so that it can more easily find rules involving more than two bonds.

We will continue working with low resolution MS data of the 150-200 monofunctional aliphatic compounds studied previously in the context of the performance program. These compounds are well-understood and thus provide a good test of the program's effectiveness. In order to insure generality in the theory formation programs, we will also test the system against the high resolution mass spectra of the 68 estrogenic steroids. Since they are also well-understood, these compounds will show how well the program can deal with complex ring systems, multifunctional compounds, cleavages involving more than two bonds, and high resolution data.

The existing programs are in good working order - within definite limits - so we expect to apply them to new sets of data from the MS laboratory as interest arises. For example, as the high and low resolution MS from marine sterols are collected we expect to use INTSUM and RULEGEN (at least) to assist in the interpretation and generalization of these data. Since these problems will advance the state of knowledge of MS, it is not correct to look on them as test problems. However, in the past the programs developed most rapidly when they were applied to unsolved problems of interest to our colleagues in the chemistry department.

For development of the interactive programs, we will rely heavily on the criteria of acceptability by Stanford users. The programs themselves will be written in INTERLISP on the SUMEX computer. Initially, we will provide interactive access to the control parameters of the programs in order to allow users to tailor their

runs to their immediate interests. Later we hope to expand these to allow interrogation of the programs with respect to both contents of the results and the program's reasoning steps.

D. Applications to Biomedical Problems

We can immediately offer to the user community the Planner, for analysis of HR/MS in terms of molecular structure. The program is insensitive to the source of the MS data, and we foresee significant use of the program for analysis of spectra of mixtures without prior separation and spectra from the GC/HRMS facility without additional programming effort. Examples of applications areas are summarized below.

We wish to exploit our existing capabilities of the analysis of biological mixtures without prior separation (ref. 33). This approach will prove particularly useful in studies of mixtures which are difficult to separate and analyze by GC. Phytoecdysones related to ecdysone, an insect molting hormone, present such a problem. GC of these compounds is very difficult, although high-pressure liquid chromatography has recently been used to carry out separations. This class of compounds represents an interesting and valuable test case for our combined MS and computer techniques, particularly the specification and subsequent acquisition of metastable defocussing data for precise linking of parent and fragment ions in the spectrum of a complex mixture (refs. 28, 33). Model compounds, mixtures and current structure elucidation problems are available (Nakanishi, Columbia; Takemoto, Tohoku University, Sendai, Japan). Although most users cannot be completely specific as to the nature of their future structure elucidation problems, we feel that several of these problems can be handled by such an approach.

As the structure generator and its extensions are developed further, we foresee continuing use of an interactive version applied to specific problems of the user community. As an example, the work in collaboration with the GRC project will involve studies of several classes of compounds extracted from human body fluids (e.g., aromatic and aliphatic acids, various classes of bases, amino acids and carbohydrates) which contain representatives varying by substitutions about a small number of molecular skeletons. The generator can define all isomers which must be considered as possible solutions.

For those problems which are amenable to attack by library search procedures, e.g., screening of GC/LRMS runs of marine sterols to weed out known compounds, we propose to use these procedures and to investigate extensions to them. using a procedure related to that described by McLafferty (K-S. Kwok, et al., J Amer. Chem. Soc., 95, 4185 (1973), we seek to determine from modified library search techniques the known structures which yield similiar spectra. Utilizing the DENDRAL structural manipulation routines, we would then seek to determine those related structures (whose spectra are not in the library) which are possible solutions. A library, including Wiswesser Line Notation names, exists (F. W. McLafferty, private communication) and would be of some utility in this work.

The MS facility in conjunction with our programs will be used in studies of the following nature:

1) Prof. Djerassi - we plan use of the MS facilities and computer programs in ongoing research connected with existing NIH-supported studies on steroids and marine sterols and continued collaboration with Prof. Adlercreutz on estrogen mixtures isolated from body fluids. Further collaboration with Prof. Adlercreutz will be on structural studies of new estrogen metabolites whose presence in mixtures has been inferred through our previous collaborative efforts.

The work on marine sterols presently utilizes GC/LRMS and frequently laborious separation procedures to isolate individual fractions for HRMS analysis. GC/HRMS will be a significant assistance in this effort. We plan MS studies of known marine sterols (utilizing INTSUM) to derive fragmentation rules, which then will be used in the Planner to aid structure elucidation of new compounds.

We also plan further work on extensions of MS theory in the steroid field, initially focussed on additional biomedically important classes of steroids related to the pregnane (progesterones) and androstane (testosterones) skeletons. This work is currently being carried out by Dr. Smith in collaboration with two visiting senior scientists (Dr. Roy Gritter, Dr. Geoff Dromey) currently on sabbatical leave fellowships.

2) Chemistry Department Collaborators - as indicated by the responses summarized in the letters of interest (Appendix A), there is significant interest in use of the MS facility by other NIH-supported members of the chemistry department. All those listed are familiar with the technique of MS as applied to structure elucidation problems. Most have used MS frequently, particularly Prof. Van Tamelen in his studies of the cyclization of squalene and related studies in the terpenoid and steroid field. The interests of these collaborators are generally in HRMS and GC/HRMS, with occasional use of other capabilities of the system. The types of compounds studied by this group and an indication of the amount of use expected are summarized in the letters of interest.

3) Genetics Research Center (GRC): (Profs. J. Lederberg, H. Cann; Dr. A. Duffield)

The body fluids analyzed by GC/LRMS to date include urine, blood, amniotic fluid and cerebrospinal fluid. Each body fluid is fractionated into the following compound classes:

a) organic acids and neutral compounds
b) amino acids
c) carbohydrates

which after appropriate derivatization are analyzed by GC/LRMS/computer system. A library of known LRMS will serve as the primary means of identifying metabolites from their experimentally recorded LRMS.

In those instances where the LRMS is insufficient for metabolite identification GC/HRMS data will be necessary to determine the composition of all ions in its mass spectrum. These data will greatly enhance the prospects of identifying the metabolite in question.

It is known (on past performance) that if a compound is present in body fluids at the level of 1 microgram per GC peak then good quality HR/MS will be recorded (ion amplitude dynamic range of 1:100, mass accuracy of +-5ppm) using the Varian MAT 711 mass spectrometer. If the GC peak of interest contains insufficient material for a HRMS scan then preparative GC could be used to concentrate that portion of the chromatograph effluent prior to GC/HRMS.

Prior to the demise of the ACME computer system (July 31, 1973) we developed a GC/HRMS system and applied it to the analysis of extracts from body fluids. The following example represents results obtained with this system during its development. The example used was a routine analysis and was run to determine the capability of the overall system during its development and not as an unknown sample of extreme interest.

The total ion plot recorded during the lifetime of the GC/HRMS analysis of an amniotic fluid is reproduced as Figure 1. A complete high resolution scan was recorded on each of the peaks shown in Figure 1. Filing time of the time-shared ACME computer system did not allow the system to operate in a repetitive scan mode. For the sake of brevity only the GC/HRMS scan (# 1594, Figure 2) corresponding to glutamic acid N-TFA O-n-Butyl ester derivative is produced. (The corresponding GC/LRMS scan is Figure 3). The scan time per decade of mass was 10.5 seconds, the resolution 6,500 and the matching tolerance for the assignment of empirical composition set to 4 mmu. The results show that the system was capable of accurate mass measurement with a dynamic range in ion amplitude of about 33:1 in this instance.

The cessation of computer support for the GC/HRMS system did not allow a HRMS analysis to be made which was crucial to the identification of a metabolite present in a body fluid. Since that time however, several instances have arisen where GC/HRMS data would have been collected in an effort to identify metabolites not previously seen.

The expected sample throughput in the GRC project with existing personnel is expected to approach 5 to 7 body fluids per week (15-21 GC/LRMS fractions to be run in the Genetics Department per week). On average GC/HRMS would be required on 1 - 2 samples per week.

The research interests of the Medical School collaborators relative to the proposed MS resource are summarized in the letters of interest (Appendix A). The MS services required by this community will include GC/LRMS (Forrest, Sera, Kalman for drug and drug metabolite identification, Rabinowitz and Wilkinson for prostaglandin identification, Robin for identification of oxidized/reduced redox pairs, Hollister for Marihuana metabolites, Barchas, neurotransmitters, Fair, polyamines and the prostatic antibacterial factor in urine); GC/HRMS (Trudell, drug metabolite identification, Kvenvolden, structure of amino acids and related compounds plus samples as required from interests described under GC/LRMS). In those instances where the biological extract contains insufficient material for a GC/HRMS scan preparative GC, using existing instrumentation within the chemistry department, can be

used to concentrate the material prior to the GC/HRMS analysis. If the material of interest is obtained relatively pure by this technique then HRMS analysis using direct sample insertion into the ion source would be utilized.

As mentioned above, several of the computer programs have immediate utility for assisting with structure elucidation problems. For example, the Structure Generator program can answer structural isomerism questions independently of mass spectrometry, (e.g. , to provide lists of isomers in conjunction with isomer interconversion problems such as carbonium ion rearrangements). Because the program will be able to generate complete lists of isomers with (or without) some specified structural features, a researcher can have confidence that no possibilities have been overlooked. Some interest in the structure generator has been expressed by representatives of the pharmaceutical industry. The generator could be used to suggest complete sets of structural alternatives for possible synthesis, once a physiologically active congener has been identified.

In more general terms, the structure generator can be richly suggestive of new, unexplored areas of synthetic organic chemistry. for example, the generator has been used by a graduate student in chemistry, Mr. Jan Simek, to identify the space of possible Diels-Alder condensation products consisting of six atoms of any combination of carbon, nitrogen, oxygen, and sulfur in a six-membered ring with one double bond. A literature search through the Ring Index revealed that many of the ring systems have never been reported.

IV. SIGNIFICANCE OF PROPOSED RESEARCH

Structure elucidation is an important and difficult problem for biomedical scientists. Many of them lack the detailed chemical background necessary to be efficient in this endeavor. Generally speaking, they also lack the frequently complex and expensive equipment (e.g., high resolution mass spectrometers) to provide spectroscopic data to assist them in solving problems of molecular structure. We plan to provide the chemical and analytical expertise to facilitate the solution of their structural problems. This research aims at providing more powerful techniques for determining molecular structures than are now routinely available. In particular, we have proposed (a) providing extended MS services as a means of collecting powerful analytic data for scientists; (b) developing (and extending) sophisticated computer programs to assist with the interpretation of the data from mass spectrometry and elsewhere, (c) developing (and extending) novel computer programs to assist with formulation of the rules of interpretation, and (d) applying these state of the art techniques to problems of biomedical relevance. Our research group is thus dedicated to a broad-based attack on the applications of structure elucidation to biological and biomedical problems.

The proposed research not only holds promise for significant long-term advances, it can have immediate benefits as well. Many members of the biomedical community at Stanford have called upon the MS laboratory for assistance in the past and will continue to do so in the future. The proposed resource will provide the conduit for a substantial increase in the utilization of MS within the Stanford biomedical community. The ability of the proposed resource to interpret the experimental data it generates (enhanced by the close proximity of the resource and biomedical community) should result in a successful program of interdisciplinary research.

HRMS is an important source of data for these problems, and GC/HRMS is still more important. Previous investment by the NIH in the Varian MAT-711 HRMS system at Stanford can be utilized now and built upon for the future. Continued operation of the GC/MS system will give the Stanford community access to state-of-the-art spectroscopic techniques and to professional mass spectroscopists who can help with ongoing problems.

The computer programs themselves constitute a unique resource for assisting with the structure determination. The previous NIH grant supported development of the programs. In part, we are requesting funds to exploit these programs.

One of the most significant aspects of this work is its interdisciplinary view of solving molecular structure problems by intelligently directed search of the space of chemical graph structures. As a result of posing the structure determination problem in this framework, we have been able to further the knowledge about structure elucidation in at least three ways. First, some of the knowledge used by analytical chemists has been made more precise for use in a computer program. Second, codifying such knowledge for the computer has led to the discovery

of new research areas to extend our existing knowledge of MS. Several publications listed in the bibliography (Refs. 42 and following) are reports of exactly this kind of research. Finally, the computer's systematic search through the space of possible structures gives the practicing scientist the confidence that no structures were merely overlooked. The efficiency of the program depends on the exclusions of many whole classes of compounds, but the computer will have rejected those classes using precise, explicitly stated criteria.

Our recent work on finding MS interpretation rules (theory formation) can provide additional unique capabilities for assisting with the problem solving. We wish to continue this research because it offers hope for a solution to the problem of furnishing real-world knowledge to computer programs -- in particular to the computer programs that assist with structure elucidation. This is a pressing problem in current AI research. High performance programs, of which DENDRAL is most often cited, derive their power from large stores of knowledge. Yet there are no routine methods for infusing such systems with knowledge of the task domain. We believe our research in theory formation holds a key to the solution of this problem.

V. FACILITIES & EQUIPMENT

The Stanford Mass Spectrometry Laboratory will provide MS services on the Varian MAT-711 mass spectrometer coupled with a Hewlett-Packard gas chromatograph (Model 7610A). As service instruments for more routine mass spectral analyses, the laboratory has a MS-9 and CH-4 mass spectrometers.

Data reduction is currently provided on Stanford's IBM 370/158 computer in conjunction with a front-end PDP-11/20 data acquisition computer. (The PDP-11/20 presently has only the capability for buffering peak profile data between the mass spectrometer and the IBM 370/158 computer at the Stanford Computer Center.) An alternative to buying time on the 370/158 is proposed and discussed in the budget justification.

The AI programs will be run on the NIH-sponsored AIM-SUMEX computer facility (a PDP-10 computer with the TENEX operating system, 192K words of memory, and adequate peripherals for our purposes). Running these programs on SUMEX will incur no charge.

VI. BIBLIOGRAPHY

A. DENDRAL PUBLICATIONS

(1) J. Lederberg, "DENDRAL-64 - A System for Computer Construction, Enumeration and Notation of Organic Molecules as Tree Structures and Cyclic Graphs", (technical reports to NASA, also available from the author and summarized in (12)).

(1a) Part I. Notational algorithm for tree structures (1964) CR.57029

(1b) Part II. Topology of cyclic graphs (1965) CR.68898

(1c) Part III. Complete chemical graphs; embedding rings in trees (1969)

(2) J. Lederberg, "Computation of Molecular Formulas for Mass Spectrometry", Holden-Day, Inc. (1964).

(3) J. Lederberg, "Topological Mapping of Organic Molecules", Proc. Nat. Acad. Sci., 53:1, January 1965, pp. 134-139.

(4) J. Lederberg, "Systematics of organic molecules, graph topology and Hamilton circuits. A general outline of the DENDRAL system." NASA CR-48899 (1965)

(5) J. Lederberg, "Hamilton Circuits of Convex Trivalent Polyhedra (up to 18 vertices), Am. Math. Monthly, May 1967.

(6) G. L. Sutherland, "DENDRAL - A Computer Program for Generating and Filtering Chemical Structures", Stanford Artificial Intelligence Project Memo No. 49, February 1967.

(7) J. Lederberg and E. A. Feigenbaum, "Mechanization of Inductive Inference in Organic Chemistry", in B. Kleinmuntz (ed) Formal Representations for Human Judgment, (Wiley, 1968) (also Stanford Artificial Intelligence Project Memo No. 54, August 1967).

(8) J. Lederberg, "Online computation of molecular formulas from mass number." NASA CR-94977 (1968)

(9) E. A. Feigenbaum and B. G. Buchanan, "Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic Chemistry", in Proceedings, Hawaii International Conference on System Sciences, B. K. Kinariwala and F. F. Kuo (eds), University of Hawaii Press, 1968.

(10) B. G. Buchanan, G. L. Sutherland, and E. A. Feigenbaum, "Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic Chemistry". In Machine Intelligence 4 (B. Meltzer and D. Michie, eds) Edinburgh University Press (1969), (also Stanford Artificial Intelligence Project Memo No. 62, July

(11) E. A. Feigenbaum, "Artificial Intelligence: Themes in the Second Decade". In Final Supplement to Proceedings of the IFIP68 International Congress, Edinburgh, August 1968 (also Stanford Artificial Intelligence Project Memo No. 67, August 1968).

(12) J. Lederberg, "Topology of Molecules", in The Mathematical Sciences - A Collection of Essays, (ed.) Committee on Support of Research in the Mathematical Sciences (COSRIMS), National Academy of Sciences - National Research Council, M.I.T. Press, (1969), pp. 37-51.

(13) G. Sutherland, "Heuristic DENDRAL: A Family of LISP Programs", Stanford Artificial Intelligence Project Memo No. 80, March 1969.

(14) J. Lederberg, G. L. Sutherland, B. G. Buchanan, E. A. Feigenbaum, A. V. Robertson, A. M. Duffield, and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference I. The Number of Possible Organic Compounds: Acyclic Structures Containing C, H, O and N". Journal of the American Chemical Society, 91:11 (May 21, 1969).

(15) A. M. Duffield, A. V. Robertson, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J. Lederberg, "Application of Artificial Intelligence for Chemical Inference II. Interpretation of Low Resolution Mass Spectra of Ketones". Journal of the American Chemical Society, 91:11 (May 21, 1969).

(16) B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, "Toward an Understanding of Information Processes of Scientific Inference in the Context of Organic Chemistry", in Machine Intelligence 5, (B. Meltzer and D. Michie, eds) Edinburgh University Press (1970), (also Stanford Artificial Intelligence Project Memo No. 99, September 1969).

(17) J. Lederberg, G. L. Sutherland, B. G. Buchanan, and E. A. Feigenbaum, "A Heuristic Program for Solving a Scientific Inference Problem: Summary of Motivation and Implementation", in R. Banerji & M.D. Mesarovic (eds.) Theoretical Approaches to Non-Numerical Problem Solving, New York: Springer-Verlag, 1970. (Also, Stanford Artificial Intelligence Project Memo No. 104, November 1969.)

(18) C. W. Churchman and B. G. Buchanan, "On the Design of Inductive Systems: Some Philosophical Problems". British Journal for the Philosophy of Science, 20 (1969), pp. 311-323.

(19) G. Schroll, A. M. Duffield, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J. Lederberg, "Application of Artificial Intelligence for Chemical Inference III. Aliphatic Ethers Diagnosed by Their Low Resolution Mass Spectra and NMR Data". Journal of the American Chemical Society, 91:26 (December 17, 1969).

(20) A. Buchs, A. M. Duffield, G. Schroll, C. Djerassi, A. B. Delfino, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J. Lederberg, "Applications of Artificial Intelligence For Chemical Inference. IV. Saturated Amines Diagnosed by Their Low Resolution Mass Spectra and Nuclear Magnetic Resonance Spectra", Journal of the American Chemical Society, 92, 6831 (1970).

(21) Y.M. Sheikh, A. Buchs, A.B. Delfino, G. Schroll, A.M. Duffield, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference V. An Approach to the

Computer Generation of Cyclic Structures. Differentiation Between All the Possible Isomeric Ketones of Composition C6H10O", Organic Mass Spectrometry, 4, 493 (1970).

(22) A. Buchs, A.B. Delfino, A.M. Duffield, C. Djerassi, B.G. Buchanan, E.A. Feigenbaum and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference VI. Approach to a General Method of Interpreting Low Resolution Mass Spectra with a Computer", Helvetica Chemica Acta, 53, 1394 (1970).

(23) E.A. Feigenbaum, B.G. Buchanan, and J. Lederberg, "On Generality and Problem Solving: A Case Study Using the DENDRAL Program". In Machine Intelligence 6 (B. Meltzer and D. Michie, eds.) Edinburgh University Press (1971). (Also Stanford Artificial Intelligence Project Memo No. 131.)

(24) A. Buchs, A.B. Delfino, C. Djerassi, A.M. Duffield, B.G. Buchanan, E.A. Feigenbaum, J. Lederberg, G. Schroll, and G.L. Sutherland, "The Application of Artificial Intelligence in the Interpretation of Low- Resolution Mass Spectra", Advances in Mass Spectrometry, 5, 314.

(25) B.G. Buchanan and J. Lederberg, "The Heuristic DENDRAL Program for Explaining Empirical Data". In proceedings of the IFIP Congress 71, Ljubljana, Yugoslavia (1971). (Also Stanford Artificial Intelligence Project Memo No. 141.)

(26) B.G. Buchanan, E.A. Feigenbaum, and J. Lederberg, "A Heuristic Programming Study of Theory Formation in Science." In proceedings of the Second International Joint Conference on Artificial Intelligence, Imperial College, London (September, 1971). (Also Stanford Artificial Intelligence Project Memo No. 145.)

(27) Buchanan, B. G., Duffield, A.M., Robertson, A.V., "An Application of Artificial Intelligence to the Interpretation of Mass Spectra", Mass Spectrometry Techniques and Appliances, Edited by G. W. A. Milne, John Wiley & Sons, Inc., 1971, p. 121-77.

(28) D.H. Smith, B.G. Buchanan, R.S. Engelmore, A.M. Duffield, A. Yeo, E.A. Feigenbaum, J. Lederberg, and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference VIII. An approach to the Computer Interpretation of the High Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids", Journal of the American Chemical Society, 94, 5962-5971 (1972).

(29) B.G. Buchanan, E.A. Feigenbaum, and N.S. Sridharan, "Heuristic Theory Formation: Data Interpretation and Rule Formation". In Machine Intelligence 7, Edinburgh University Press (1972).

(30) Lederberg, J., "Rapid Calculation of Molecular Formulas from Mass Values". Jnl. of Chemical Education, 49, 613 (1972).

(31) Brown, H., Masinter L., Hjelmeland, L., "Constructive Graph Labeling Using Double Cosets". Discrete Mathematics (in press). (Also Computer Science Memo 318, 1972).

(32) B. G. Buchanan, Review of Hubert Dreyfus' "What Computers Can't Do: A Critique of Artificial Reason", Computing Reviews (January, 1973). (Also Stanford Artificial Intelligence Project Memo No.

(33) D. H. Smith, B. G. Buchanan, R. S. Engelmore, H. Aldercreutz and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference IX. Analysis of Mixtures Without Prior Separation as Illustrated for Estrogens". Journal of the American Chemical Society 95, 6078 (1973).

(34) D. H. Smith, B. G. Buchanan, W. C. White, E. A. Feigenbaum, C. Djerassi and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference X. Intsum. A Data Interpretation Program as Applied to the Collected Mass Spectra of Estrogenic Steroids". Tetrahedron, 29, 3117 (1973).

(35) B. G. Buchanan and N. S. Sridharan, "Rule Formation on Non-Homogeneous Classes of Objects". In proceedings of the Third International Joint Conference on Artificial Intelligence (Stanford, California, August, 1973). (Also Stanford Artificial Intelligence Project Memo No. 215.)

(36) D. Michie and B.G. Buchanan, "Current Status of the Heuristic DENDRAL Program for Applying Artificial Intelligence to the Interpretation of Mass Spectra". August, 1973.

(37) H. Brown and L. Masinter, "An Algorithm for the Construction of the Graphs of Organic Molecules", Discrete Mathematics (in press). Also Stanford Computer Science Department Memo STAN-CS-73-361, May, 1973)

(38) D.H. Smith, L.M. Masinter and N.S. Sridharan, "Heuristic DENDRAL: Analysis of Molecular Structure," Proceedings of the NATO/CNA Advanced Study Institute on Computer Representation and Manipulation of Chemical Information, in press.

(39) R. Carhart and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference XI: The Analysis of C13 NMR Data for Structure Elucidation of Acyclic Amines", J. Chem. Soc. (Perkin II), 1753 (1973).

(40) L. Masinter, N. Sridharan, and D.H. Smith, "Applications of Artificial Intelligence for Chemical Inference XII: Exhaustive Generation of Cyclic and Acyclic Isomers.", submitted to Journal of the American Chemical Society.

(41) L. Masinter, N. Sridharan, R. Carhart and D.H. Smith, "Applications of Artificial Intelligence for Chemical Inference XIII: An Algorithm for Labelling Chemical Graphs", submitted to Journal of the American Chemical Society.

(42) The Determination of Phenylalanine in Serum by Mass Fragmentography. Clinical Biochem., 6 (1973). By W.E. Pereira, V.A. Bacon, Y. Hoyano, R. Summons and A.M. Duffield.

(43) The Simultaneous Quantitation of Ten Amino Acids in Soil Extracts by Mass Fragmentography. Anal. Biochem., 55, 236 (1973). By W.E. Pereira, Y. Hoyano, W.E. Reynolds, R.E. Summons and A.M. Duffield.

(44) An Analysis of Twelve Amino Acids in Biological Fluids by Mass Fragmentography. Anal. Chem., in press. By R.E. Summons, W.E.

Pereira, W.E. Reynolds, T.C. Rindfleisch and A.M. Duffield.

(45) The Quantitation of B-Aminoisobutyric Acid in Urine by Mass Fragmentography. Clin. Chim. Acta, in press. By W.E. Pereira, R.E. Summons, W.E. Reynolds, T.C. Rindfleisch and A.M. Duffield.

(46) The Determination of Ethanol in Blood and Urine by Mass Fragmentography. Clin. Chim. Acta, in press. By W.E. Pereira, R.E. Summons, T.C. Rindfleisch and A.M. Duffield.

Publications Describing DENDRAL-Related Research But Not Funded By This Grant

(47) An Automated Gas Chromatographic Analysis of Phenylalanine in Serum. Clinical Biochem., 5, 166 (1972). By E. Steed, W. Pereira, B. Halpern, M. D. Solomon and A.M. Duffield.

(48) Pyrrolizidine Alkaloids. XIX. Structure of the Alkaloid Erucifoline. Coll. Czech. Chem. Commun., 37, 4112 (1972). By P. Sedmera, A. Klasek, A.M. Duffield and F. Santavy.

(49) Chlorination Studies I. The Reaction of Aqueous Hypochlorous Acid with Cytosine. Biochem. Biophys. Res. Commun., 48, 880 (1972). By W. Patton, V. Bacon, A.M. Duffield, R. Halpern, Y. Hoyano, W. Pereira and J. Lederberg.

(50) A Study of the Electron Impact Fragmentation of Promazine Sulphoxide and Promazine using Specifically Deuterated Analogues. Austral. J. Chem., 26, 325 (1973). By M.D. Solomon, R. Summons, W. Pereira and A.M. Duffield.

(51) Spectrometrie de Masse VIII. Elimination d'can Induite par Impact Electronique dans Le Tetrahydro-1,2,3,4-Napthtal-ene-diol-1,2. Org. Mass Spectre., 7, 357 (1973). By P. Perros, J.P. Morizur, J. Kossanyi and A.M. Duffield.

(52) Chlorination Studies II. The Reaction of Aqueous Hypochlorous Acid with a-Amino Acids and Dipeptides. Biochim. et Biophys. Acta, 313, 170 (1973). By W.E. Pereira, Y. Hoyano, R. Summons, V.A. Bacon and A.M. Duffield.

(53) Spectrometrie de Masse. IX. Fragmentations Induites par Impact Electronique de Glycols- En Serie Tetraline. Bull. Chim. Soc. France, 2105 (1973). By P. Perros, J.P. Morizur, J. Kossanyi and A.M. Duffield.

(54) The Use of Mass Spectrometry for the Identification of Metabolites of Phenothiazines. Proceedings of the Third International Symposium on Phenothiazines, Raven Press, New York, 1973. By A.M. Duffield.

(55) Chlorination Studies IV. The Reaction of Aqueous Hypochlorous Acid with Pyrimidine and Purine Bases. Biochem. Biophys. Res. Commun., 53, 1195 (1973). By Y. Hoyano, V. Bacon, R.E. Summons, W.E. Pereira, B. Halpern and A.M. Duffield.

(56) Mass Spectrometry in Structural and Stereochemical Problems. CCXXXVII. Electron Impact Induced Hydrogen Losses and Migrations in Some Aromatic Amides. Org. Mass Spectry., in press. By A.M. Duffield, G. deMartino and C. Djerassi.

(57) Stable Isotope Mass Fragmentography: Quantitation and Hydrogen-Deuterium Exchange Studies of Eight Murchison Meteorite Amino Acids. Geochem. et Cosmochim. Acta, submitted for publication. By W.E. Pereira, R.E. Summons, T.C. Rindfleisch,